Social Work & Social Sciences Review 12(2) 2005 pp.5-28

The elusive search for the silver bullet in prevention and family support programs for vulnerable families

Terry Carrilio[1]

***Abstract**: Since the 1970s home visiting has been seen as a promising prevention and family support strategy. Programmes proliferated, with various funding, conceptual, and structural characteristics. Policy makers, funders and practitioners have been avidly seeking evidence of programme effectiveness. Research results have been disappointing, often showing modest results. The search for clarity continues, often with high political and programmatic stakes. This paper describes the results from three projects, representing elaborations of the Healthy Families model, covering 25 sites. Findings suggest that programme effects were experienced differentially based on ethnicity and initial mental health status.*

***Keywords**: home visiting, prevention. family support, programme evaluation*

Introduction

Over the past two decades home visiting for the purpose of supporting vulnerable families has gained considerable attention by policy makers, practitioners and researchers. Family support programmes are multi-service interventions aimed at enhancing family resiliency and preventing adverse outcomes for vulnerable families (Carrilio, 1998; Shonkoff & Phillips, 2000; Carrilio, 2001; Guterman, 2001; Hall et al., 2002; Duggan et.al., 2004). Recent data on brain development and the importance of the first three years of a child's life have led to a proliferation of efforts to enhance the well being and overall functioning of overburdened families (Kotulack, 1995; Garbarino, 1995; Shonkoff, & Phillips, 2000; Karoly, et al., 1998; Reynolds et al., 2001; Daro & Cohn-Donnelly, 2002). Recognition of the potentially devastating consequences of early developmental traumas and deprivations (Van der Kolk, et al., 1994; Kernberg, 1999; Erickson & Kurtz-Reimer, 1999; Barlow, 2003) has led

1. Assistant Professor.
Address for correspondence: San Diego State University School of Social Work, 5500 Campanile Drive, San Diego, CA 92182-4119, USA. tbear10009@aol.com.

to the development of programmes that are aimed broadly at helping families build their strengths and provide effective launching platforms for successful children (Garbarino, 1995; Carrilio, 2001; Karoly, et.al., 1999; Shonkoff,, & Phillips, 2000; Daro, 2005).

Evaluations of these programmes have often been high stakes propositions (Sweet, & Appelbaum, 2004; Sherwood, 2005; Hahn, et al., 2005), with policy, funding and programme structure hanging in the balance. The gold standard of randomized clinical studies was established early on, and although there are researchers suggesting alternative research paradigms (Daro, 2005; Hahn, et.al., 2005), home visiting programmes continue to be reviewed in a high pressure environment with high expectations. The proverbial 'silver bullet'—the intervention or combination of interventions that will improve family functioning and reduce adverse outcomes has yet to be identified. The findings emerging from evaluations of home visiting programmes are conflicting and difficult to interpret (Daro & Cohn-Donnelly, 2002; Sweet, & Appelbaum, 2004; Duggan et.al., 2004; Hahn et.al., 2005).

There is increasing evidence that prevention and broad based support for families does lead to positive results (Karoly, et.al., 1998; Sweet & Appelbaum, 2004; Hahn, et.al., 2005), although the results are often difficult to understand and specific programme characteristics leading to effectiveness do not emerge consistently from the studies (Sweet & Appelbaum, 2004). As research in the field has become more refined, the discussions have moved from looking for the 'best' model to understanding characteristics of families and helpers that contribute to programme effectiveness (Olds, 2003; Wagner, 2003; Daro, et.al., 2003). Additionally, issues of model fidelity and implementation context within organizations and communities have emerged as important factors in understanding programme effectiveness (McGuigan, et.al., 2003). There is increased awareness that it is important to consider the funding and policy pressures affecting programmes as well as understanding the differences in community contexts within which the families and communities are functioning (McGuigan, et.al., 2003; Carrilio, et.al., 2003; Carrilio, 2003)

This paper describes the results of three projects funded by the California Department of Social Services over a seven year period (Carrilio, 1998; Landsverk, et.al., 2001; Carrilio, et.al., 2002; Carrilio, & Min, 2003). The programmes shared many common elements and were all intended to build upon the Healthy Families model, but were funded and structured differently.

The three programmes

The projects were:

1. Healthy Families-San Diego (HFSD), a randomized clinical trial taking place at one site (1995-2000);
2. California Safe and Healthy Families (Cal-SAHF), a quasi-experimental study with 7 sites throughout California (1998-2001);
3. Answers Benefiting Children (ABC), an action research (Patton, 1997; Whyte, 1991) project with 17 sites throughout the state of California (1999-2002).

Table 1 outlines key elements of the three programmes and highlights the outcomes measures that were shared, as well as the research strategies utilized in the programme evaluations. All three programmes collected similar process data, utilizing versions of the same MIS.

Method and rationale for the current study

The three projects were initiated and evaluated separately. The results are summarized in Table 2. While the research results, especially on HFSD did not show clear intervention effects, the positive trends were intriguing. Additional analysis was conducted with the following questions in mind:

- Did some clients improve?
- What are the demographic and initial risk characteristics of those who improved and those who did not improve?
- When we have similar clients with respect to risk and demographics, does the service delivery pattern make a difference in terms of who does better on the outcomes and risk improvement variables?
- Can we identify service delivery patterns (combination of intensity, duration, continuity, who delivered) that predict better outcomes?
- Does implementation affect client outcomes?

The analysis consisted of a descriptive comparison of the three programmes with respect to demographic, risk and outcomes variables. A comparison was conducted of the services actually received by families in each programme. Further analysis, using hierarchical linear modeling (HLM) was conducted on the HFSD and CAL-SAHF samples (Raudenbush, & Bryk, 2002).

Table 1
Comparison of the key elements in the three programme initiatives

Program Element	HFSD	Cal-SAHF	ABC
Target Population	Used a 2-stage screening process in-hospital. All clients at risk cut point offered randomization. Child Protection Service (cps) clients excluded	Community defined criteria. Clients could self refer. At risk clients, but not cps cases. Each program defined risk criteria	At risk families defined by each site
Duration Of Services	36 months	24 months	36 months
Caseload Size	20-25	20-25	20-25
Staff Characteristics	*Team leader:* licensed MSW *Home visitors:* BA preferred, some AA's accepted 2 of 10 were males *Specialists*: (child development & group)—masters level	*Team leader:* licensed master's level. In some cases a BA person was used *Home visitors:* B.A. preferred, some had only High School Diploma (H.S.) *Specialists*: (nurse, substance abuse, group, child development) masters' level preferred, except nurse, who was required to be an R.N.	*Team leader:* licensed master's level. In some cases a BA person was used *Home visitors:* B.A. preferred, some had only H.S. Use of Americorps workers *Specialists*: (nurse, substance abuse, group, child development) masters' level preferred, except nurse, who was required to be an R.N.
Home Visiting	Weekly at first–tapers over time	Weekly at first–tapers over time	Weekly at first–tapers over time
Groups	Not required but suggested. Parent groups and age appropriate children's groups were offered.	Weekly integrated didactic/support groups for parents. Children's groups—age-specific	Weekly integrated didactic/support groups for parents Children's groups—age-specific
Training	40 hour training. On-going monthly training	40 hour training. On-going monthly training	40 hour training. On-going monthly training
Multiple Sites	No	7 Sites	17 Sites
Funding	Three year grant funding	Capitated contract—reimbursed for services delivered	Programme contract

Research Strategy	Randomized intervention (247) and control (241) Independent research interviews at baseline, year1,year2,year3	Randomized intervention(248), using HFSD control group as a control Independent research interviews at baseline, year1,year2,year3	Action research, utilizing field surveys, interviews & feedback meetings at year 1, 2, 3.
Outcomes Domains & Measures[1]	1. Maternal well-being (*CES-D,Rand Mental Health Inventory; demographic markers) 2. Family violence(*CTS2, PC-CTS) 3. Family Functioning (Wasik Problem-Solving Scale (Wasik et.al.,1997), *PSI,NCAST,CLSS) 4. Substance Use (*CAGE) 5. Child Health (Medicaid MIS download; health visits; CPS reports) 6. Child development (*HOME; *Bayley MDI;NCAST; CBCL; Stanford-Binet) 7. School readiness (Zimmerman Pre-School Language Scale; Vineland Adaptive Behavior Scales) 8. Service delivery (process data--frequency of visits, annual home visit rates, content of home visits, service levels, caseload sizes, group attendance, contact with multiple team members)	1. Maternal Well-Being (*CES-D,Rand Mental Health Inventory; demographic markers) 2. Family violence(*CTS2, PC-CTS) 3. Family Functioning (Wasik Problem-Solving Scale (Wasik et.al.,1997), *PSI,NCAST,CLSS) 4. Substance Use (*CAGE) 5. CPS reports) 6. Child development (*HOME; *Bayley MDI;NCAST; CBCL; Stanford-Binet) 7. Service delivery (process data--frequency of visits, annual home visit rates, content of home visits, service levels, caseload sizes, group attendance, contact with multiple team members)	1. Maternal well being (demographic markers, *CES-D) 2. Family and parental functioning (AAPI-2 (Bavoleck & Keene,1999), *CTS2) 3. Substance use (DAST(Gavin et.al., 1989).AUDIT (Saunders,J.B., et.al.,1993)) 4. Service delivery (process data-frequency of visits, annual home visit rates, content of home visits, service levels, caseload sizes, group attendance, contact with multiple team members) 5. Organizational context (interviews, surveys, TEP, budget analysis, facilities analysis, staff retention review 6. Community stakeholder interviews, surveys, context analysis of planning documents) 7. Implementation (8 domains compared across sites), focus groups, parent surveys, case file reviews)

1. Measures shared by one or more of the programs are asterisked. The following summarizes the measures that were shared among the programs and included in the current analysis:

CES-D (Center for Epidemiologic Studies Depression Scale): Used by all three programs
CAGE (4-item alcohol screening tool): Used by HFSD and Cal-SAHF CTS (Conflict Tactics Scale): Used by all three programs
HOME (Home Observation for Measurement of the Environment): Used by HFSD and Cal-SAHF
PSI (Parenting Stress Index) Used by HFSD and Cal-SAHF Bayley MDI (Mental Development Index): Used by HFSD and Cal-SAHF

Table 2
Summary of research characteristics and results for HFSD, Cal-SAHF, and ABC

HFSD

- Carried out from February, 1996 through March, 1997
- 488 families at high risk for child abuse and neglect were selected through a two stage process at the time of the index child's birth at Mary Birch Hospital and enrolled in the clinical trial (247 intervention families and 241 control families).
- Three years of services were provided for the 247 intervention families.
- Interviews and assessments were completed on the 488 clinical trial families (including the 241 control group families) at baseline, year 1, year 2, and year 3 with retention rates of 89%, 83% and 85% respectively.
- The randomization was successful with no statistically significant differences (except for enrollment in managed care Medical) observed between the intervention and control groups at baseline.
- No differential attrition between the two groups has been observed
- The original proposal for the study specified the service delivery components to include the following service elements: (1) in-home supportive services; (2) support groups and parenting classes; and (3) case management
- Funding–once the sample was identified, no new clients were added
- Data collection through MIS, research interview

Some promising trends:

- Positive outcomes have been observed in the areas of child preventive health care
- Reduction in maternal depressive symptoms
- Reduced psychological aggression by mother toward index child
- Positive trends in the developmental functioning of the index child
- Intervention families reported significantly more well child visits and indicated a significantly higher proportion with well child visit compliance than did control families
- Intervention mothers more likely to be in school

CAL-SAHF

- Carried out from March,1998-February 2000
- 7 sites throughout California
- 36 families from each site randomly selected as a research intervention group
- Interviews and assessments were completed on the 252 intervention families (including the 241 control group families) at baseline, year 1, year 2, and year 3
- the HFSD control group was used as the Cal-SAHF control group
- Two years of services were provided for the 252 intervention families (there were more families in the initiative, but this analysis includes only those who were randomized and received the same baseline, yr 1,yr2,yr3 research interviews as the HFSD controls)
- Service delivery was an enhanced version of the HFSD program
- Service funding-program was capitated

There were *positive trends* on the same dimensions as seen in HFSD:

- Improvements in well-baby care and preventive health care
- Reduced psychological aggression by mother toward index child
- Positive trends in the developmental functioning of the index child
- Intervention mothers more likely to be in school

Additional positive trends:

- Reduced maternal substance abuse
- Reduced use of corporal punishment

ABC

- Carried out originally from July, 1999 through June 2002
- 17 sites throughout California
- Integrated the Cal-SAHF program model into Family Resource Centers and incorporated a requirement for county level collaboration and planning
- No random selection
- Specific focus on systems change
- Funding-through the counties, with an expectation that planning for sustainability would begin immediately with project start

The data analysis identified some key issues associated with implementation, including: readiness, leadership, quality, and funding.

Positive trends:

- Improvement on all five domains of the Adult and Adolescent Parenting Inventory (AAPI-2)
- Reduction in parental depression scores on the CES-D
- Significant improvement in scores on the Maternal Social Support Index
- Reduction in Moderate Physical Aggression as measured by the Conflict Tactics Scale (CTS)

Results

Table 3
Selected Demographic Characteristics in the Three Samples

Demographic characteristics	HFSD Control (*N* = 241)	HFSD Intervention (*N* = 248)	Cal-SAHF (*N* = 247)	ABC (*N* = 7556)
Mean age at baseline	23.8	23.3	23.5	29.4
SD	6.1	6.1	6.5	9.2
	%	%	%	%
Marital status				
Single	77.5	77.9	63.6	28.3
Married	15.0	14.3	30.5	42.0
Divorced	5.0	4.9	4.7	7.9
Separated	2.1	2.5	1.3	9.4
Others	0.4	0.4	0.0	12.4
Educational Category				
No HS Diploma	53.1	55.5	64.1	50.5
HS/GED Grad	19.9	23.1	21.8	27.2
Some college	27.0	21.5	14.15	22.3
Ethnicity				
Hispanic	43.6	48.6	62.1	55.9
White	25.7	22.7	15.7	33.5
African American	19.5	19.4	16.1	4.8
Asian/Other	11.2	9.3	6.0	5.9
Language Spoken				
English	51.0	49.0	67.0	58.0
Spanish	40.0	44.0	32.0	40.5
Others including Asian language	9.0	7.0	1.0	1.5

Table 3 compares the three programmes with respect to demographic variables, Table 4 compares the three programmes with respect to risk factors, and Table 5 compares selected outcomes variables.

The average age of participants in HFSD and Cal-SAHF projects was approximately 24 years, whereas ABC participants were relatively older at 29 years. Three-quarters of the HFSD participants were single at the time of baseline. About 64 percent of the Cal-SAHF participants were single. By way of contrast, only 28 percent of the participants in the ABC project were single and about 42 percent were married. More than half of the participants from the three programmes did not complete high school. Hispanic

participants comprised the majority of the participants across the programmes, ranging from 43% to 62%. The most notable differences among the three study populations are in the areas of age and marital status. The three populations are similar with respect to ethnicity, language spoken, and education (see table 4 below).

At baseline, almost half of the participants from all of three projects fell into the clinically depressed range on the CES-D (Radloff, 1977). Average CES-D scores were marginally greater than the clinical cutoff point of 16 points. The proportion of those in the clinically depressed group ranged from 42.7% for Cal-SAHF to 53% for the intervention group of the HFSD project. Although the proportion of those who score in the clinically depressed group decreased from baseline to the end of each project, it is noteworthy that a substantial proportion of participants in each of the projects continued to evidence depression at the end of the project intervention. It appears that in all three samples there was a decrease in substance use during the programme, and in the ABC programme the reduction in substance use was statistically significant.

Table 4 indicates that as each project progressed, the number of average incidences of moderate aggression behaviors as measured on the CTS (Strauss, et.al., 1996) increased. The same trend was observed for reported incidents of severe aggression. These findings are troublesome in that the reduction of violence in the home was one of the key risk factors that these projects were intended to address. Two explanations for the increase in reported instances of moderate and severe aggression should be considered: 1) as families

Table 4
Comparison of risk factors in the three samples

Variable	HFSD Control (*N* = 241)	HFSD Intervention (*N* = 248)	Cal-SAHF (*N* = 247)	ABC (*N* = 7556)
Clinically depressed				
(16 or more CES-D score)				
Baseline	45.2%	52.6%	42.7%	50.6%
Year 1	41.8%	39.4%	37.1%	-
Year 2	37.6%	37.2%	26.4%	-
Year 3	28.8%	32.8%	-	41.3%*** (*n* = 982)
Family violence				
Moderate aggression				
Baseline	-	-	-	1.81*+
Year 1	53.8%	50.8%	43.0%	-
Year 2	72.7%	75.6%	67.4%	-
Year 3	77.7%	69.7%	-	1.6* (*n* = 930)

Severe aggression				
Baseline	-	-	-	.88*+
Year 1	4.5%	2.1%	1.7%	-
Year 2	10.9%	7.9%	4.1%	-
Year 3	9.6%	5.3%	-	.56***
				(*n* = 933)
Substance use - Mom				
Baseline	-	-	-	-
Year 1	17.7%	21.4%	-	-
Year 2	14.0%	14.8%	-	-
Year 3	10.2%	8.5%	-	-
CAGE (0-4)				
Baseline	.71	.72	.60	-
Year 1	.33	.34	.15	-
Year 2	.24	.13	.11	-
Year 3	.12	.12	-	-
CAGE				
(2 or more positive responses)				
Baseline	19.1%	19.9%	17.4%	-
Year 1	9.5%	10.2%	4.2%	-
Year 2	6.8%	4.1%	2.5%	-
Year 3	2.9%	4.0%		
AUDIT				
First Assessment	-	-	-	1.4*+
Last Assessment	-	-	-	1.19***
				(*n* = 944)
DAST				
First Assessment	-	-	-	.95*+
Last Assessment	-	-	-	.78***
				(*n* = 897)

* CAL-SAHF project only provided two years of services

** Not all clients received a baseline and final assessment

*+ Data were reported in terms of average score

*** Significant differences between first and last assessment at $p < .05$

became more comfortable with the programme they were more likely to accurately report troublesome behaviors; and, 2) the child's development and attainment of some autonomy may provoke reactions from stressed parents (see Table 5 below).

The Bayley Mental Development Index (Bayley, 1993) was used in more than one of the programmes to track child development outcomes. For HFSD and Cal-SAHF, the average Mental Development Index (MDI) scores decreased from Year 1 to Year 2 by about 10 points. In both of the programmes, the percentage of children with an MDI score of 85 or lower increased considerably. However, the rate of decrease in MDI scores was significantly lower for the intervention groups. Nevertheless, the finding that developmental scores were decreasing, even with the intervention, is disturbing. One explanation for our findings is that the vulnerabilities of the families are such that even though the intervention was able to reduce the rate of decline in development, in order to halt or even reverse the decline, it may be necessary to enhance specific programming aimed at improving child developmental outcomes.

The Home Observation for the Measurement of the Environment (HOME) Inventory assesses parenting practices using the following subscales: acceptance of child's behavior, opportunity for stimulation, organization of the environment, parental involvement, parental responsivity, and appropriate play materials (Caldwell, & Bradley, 1984). The HOME is based on parent reports and field evaluator observations. While there is some improvement in the HOME scores for both the HFSD and Cal-SAHF samples, the change is not significant, nor is there a significant difference between the HOME scores for the control and the intervention groups in HFSD.

Table 5
Comparison of risk factors in the three samples

Selected outcome variables (mean scores)	HFSD Control (*N* = 241)	HFSD Intervention (*N* = 248)	Cal-SAHF (*N* = 247)	ABC (*N* = 7556)
Bayley Mental Development Index (MDI)[1]				
Baseline	-	-	-	-
Year 1	102.5	105.0	96.4	-
Year 2	89.9	92.5	89.4	
Year 3	-	-	-	-
Proportion of Bayley MDI < 85				
Baseline	-	-	-	-
Year 1	5.6%	4.3%	8.2%	-
Year 2	30.9%	22.2%	33.8%	-
Year 3	-	-	-	-

HOME Inventory[2]				
Baseline	-	-	-	-
Year 1	35.2	35.6	34.6	-
Year 2	34.6	34.6	34.9	-
Year 3	40.2	40.1	-	-
Parenting Stress Index (PSI)[3]				
Baseline	-	-	-	-
Year 1	71.2	72.5	77.8	-
Year 2	74.6	72.8	77.0	
Year 3	72.8	71.2		
AAPI-2 Subscale A[1]				
Baseline	-	-	-	20.0
Last Assessment	-	-	-	21.1***
AAPI-2 Subscale B[1]				
Baseline	-	-	-	36.4
Last Assessment	-	-	-	38.1***
AAPI-2 Subscale C[1]				
Baseline	-	-	-	38.6
Last Assessment	-	-	-	39.9***
AAPI-2 Subscale D[1]				
Baseline	-	-	-	21.7
Last Assessment	-	-	-	23.7***
AAPI-2 Subscale E[1]				
Baseline	-	-	-	19.1
Last Assessment	-	-	-	19.6***

1 *Bayley Mental Development Index (MDI)* This instrument provides overall standardized scores for infant development in the areas of mental and motor development. There is also a behavior rating scale. The mean Mental Development Index (MDI) score is 99.8., with a standard deviation of 14.9.

2. *Home Observation for Measurement of the Environment (HOME)* This is an instrument with six subscales, intended to measure the home environments of infants and toddlers. It is an observational measure, with a simple "yes/no response to indicate the presence or absence of a factor. The higher the score, the more the home environment is considered conducive to optimal child development. The instrument is normed for infants through children age 10.

3. *Parenting Stress Index (PSI)* The short form of this Parenting Stress Index consists of 36 items that measure parent-child interactions. There are five subscales. The instrument relies on self report and identifies areas in which parent child interactions may be dysfunctional. Higher scores on this instrument reflect higher levels of stress.

The three programmes all required weekly home visits over an extended period of time. Table 6 shows a comparison among the programmes on some of the key variables of service delivery.

Table 6
Comparison of service provision in the three programme

	HFSD Intervention (*N* = 247)	Cal-SAHF (*N* = 248)	ABC (*N* = 10282)
Expected duration in programme (from intake to termination)	36 months	24 months	24 months
Average duration in programme as % of expected duration	17 months 47%	12.6 months 51%	5.8 months 24%
Types of Services			
Case management	63.2%	22.1%	28.5%
Home visits	24.1 %	66.6%	39.2%
Group services	0.1%	1.8%	5.6%
Crisis intervention	0.1%	0.1%	1.4%
Concrete services	1.7%	4.0%	4.6%
Health education	10.8%	5.5%	2.3%
Other	-	4.9%	18.4%
Location of Services			
Programme Office	3.7%	7.2%	35.9%
In-Home	47.3%	72.4%	31.6%
Telephone	45.5%	16.8%	25.2%
Others	3.4%	3.6%	7.3%
Recipient of Services			
Mother	69.2%	82.6%	63.0%
FOB or Male Partner	2.3%	1.1%	7.0%
Mother and Male Partner	4.9%	1.7%	17.0%
Others	23.6%	7.9%	13.0%

Although the programme model for all three projects was similar, the service delivery patterns appear to be quite different. It is noteworthy that HFSD was the most rigorous of the projects, and that client retention was high in that programme. Cal-SAHF may have shown the highest retention because the programme was captitated and there were strong incentives to retain clients and provide reimbursable services. This is also reflected in the larger percentage of reported home visits, since programmes were primarily reimbursed for home visits provided. Clearly, in ABC,

as control over model fidelity was attenuated by 'going to scale' in seventeen sites, the level of client retention-was reduced, and service delivery seems to have been reduced. It is noteworthy that in the ABC project, where an effort was made to better integrate centre-based and home based services, the location of service delivery differed markedly from HFSD and Cal-SAHF. In ABC more men received services. This may be a function of the demographic makeup of the ABC population (more married couples) and specific outreach to men, which was encouraged by the funder.

Hierarchical linear modelling (HLM) and the HFSD and CAL-SAHF samples

The HFSD sample represents the most rigorous of the programmes, and while there are positive trends in the data, the programme effects do not approach significance on most of the outcomes measures (Gomby, 1999; Lansdverk, et.al., 2001; Sweet, & Appelbaum, 2004; Sherwood, 2005). The ABC programme was designed to utilize the HFSD control group as a proxy control group, and the results were consistent with the HFSD data (Carrilio, et.al. 2002; Carrilio, & Min, 2003). However, with the ABC sample which followed a quasi experimental design, significant differences were identified between baseline and end of programme measures on key outcome variables (Carrilio, et.al. 2002). Some interesting findings, particularly with respect to the impact of ethnicity and depression emerged in HFSD (Landsverk, et al, 2001) and Cal-SAHF. For this reason, a more detailed HLM analysis of the HFSD and Cal-SAHF data sets were undertaken in an effort to better identify factors associated with programme effects. Results are reported below in tables 7, 8 and 9.

Table 7 indicates that in the HFSD and Cal-SAHF populations, there was a significant decrease in maternal depression that was not directly attributable to the intervention. The HFSD group includes both intervention and control subjects in order to observe individual change over time. Age, marital status, and assignment to the intervention group did not seem to affect depression scores. Interestingly, when ethnic differences were considered in the HFSD population, it could be seen that Latino/Spanish speaking clients had lower depression scores than base status, at both baseline and year three, while in Cal-SAHF the Latino/Spanish speaking clients showed higher depression scores than base status, at both time points. Also noteworthy is the observation that Asian women in the HFSD population demonstrated higher depression scores than base status, at both baseline and year three. In HFSD low baseline income was found to be significantly related to depression scores at baseline but not at year three. Education to high school level (ie beyond 8th grade[1]) in HFSD was related to lower depression scores both at baseline and at year three. The rates of reduction in depression in the HFSD intervention group were more rapid than for the control group members. In Cal-SAHF, the rate of reduction in depression was

fairly consistent for all participants. Also noteworthy, in the HFSD population was the observation that there is a positive correlation between the number of services received and depression, indicating that those who were depressed appear to have received a higher number of service transactions.

Table 7
Summary of HLM findings on depression -HFSD and Cal-SAHF programmes

	HFSD (n = 339)		Cal-SAHF (n = 200)	
Depression (CES-D)	Baseline	Year 3	Baseline	Year 2
Initial status				
Base Status [1,3]	17.0*--	14.8*	14.9*	11.4*
Age	-	-	-	-
Latino/Spanish Speaking	lower[2] *	lower*	higher*	higher*
Asian	higher[2] *	higher*		
Baseline Income	lower*	-		
Intervention	-	-		
Education beyond 8th Grade	lower*	lower*		
Married	-	-		
Growth rate				
Base rate	-1,12*			

* denotes statistical significance at .10 or less

1. Base Status of HFSD is different from that of Cal-SAHF. For HFSD, it denotes a base status value of the depression score at the first follow-up (i.e., an intercept or constant), whereas for Cal-SAHF, the base status represents the depression score for those who are Latino-Spanish speaking and have high school education (beyond 8th grade).
2. The term 'lower' or 'higher' indicates that expected score for each predictor (i.e., Latino/Spanish speaking, Asian, Baseline income, etc.) is either lower or higher than Base Status reported at the top of each column.
3. In HLM a base status, representing the mean score of a group, using combined variables, is used to compare the scores of different groups. The base status is determined statistically, using theoretical expectations and groupings in which there is sufficient data to meet statistical requirements. The expected scores for other groupings on the identified variables are then compared to see if the expected values of each differ significantly from the base status.

Table 8 indicates that in HFSD, there was a decline in parental stress scores from baseline (81.3) to Year Three (79.7). In Cal-SAHF, however, Parental Stress Index (PSI)scores (Abidin, 1995) increased. In HFSD, age was associated with a lower PSI score at baseline, but not at year 3. As for the relationship between parental stress and ethnicity, in HFSD, Latino/Spanish speaking parents were experiencing more stress than base status at baseline and achieved a significant reduction by year 3. In Cal-SAHF, the Latino/Spanish speaking cohort did not show significant differences from the rest of the sample on PSI scores at baseline, but they did show significantly lower scores at the end of year 2. While white clients in HFSD did not show significant differences from the population on PSI scores, in Cal-SAHF, white parents had significantly lower PSI scores both at baseline and at 2 years. Being assigned to the intervention group did not seem to affect PSI scores in HFSD. However, at baseline, parents with higher educations were exhibiting more stress. Marital status did not affect PSI scores. Noteworthy, however, is a significant effect of chronic depression in both Cal-SAHF and HFSD: in both populations, those with chronic depression reported significantly higher level of parenting stress both at baseline and at the end of services. While baseline income did not show significant changes from baseline to the end of services in both HFSD and Cal-SAHF, in HFSD there was a significant relationship population wide between low incomes and elevated parenting stress scores.

Table 9 indicates that in HFSD, there were statistically significant improvements in HOME scores from baseline to the end of services. For the HFSD population, HOME scores were higher at baseline for parents who were older than base status. However, the reverse was true of the Cal-SAHF population, with lower HOME scores associated with older parents. In the Cal-SAHF programme, Latino English speaking clients showed significantly lower HOME scores at baseline, but by the end of the project they did not differ significantly from the rest of the sample. For the HFSD sample, Latino/Spanish speaking families consistently demonstrated lower HOME scores, but this was not the case for the Cal-SAHF population. In HFSD, African American families received lower HOME scores at baseline. For HFSD, while Asian families started out very similar to the rest of the population, at the end of year 3, they showed significantly higher HOME scores. In Cal-SAHF, the Asian families started with lower HOME scores and did not differ from the overall study population by the end of the project. In HFSD, higher education was associated with higher HOME scores. In Cal-SAHF this was not observed. Likewise, with respect to chronic depression, there was a relationship in the HFSD sample between depression and lower HOME scores both at baseline and year 3.

Table 8
Summary of HLM findings on Parenting Stress Index (PSI)

HFSD	Cal-SAHF (n = 338)		(n = 120)	
Parenting Stress Index (PSI)	Baseline	Year 3	Year 1	Year 2
Initial Status				
Base Status [1,3]	81.3*	79.7*	81.3*	82.4*
Age	lower[2] *	-	-	-
Latino/English Speaking	-	-	-	lower *
Latino/Spanish Speaking	higher[2] *	lower	-	*lower *
Asian	-	-	-	-
White	-	-	lower *	lower *
Intervention	-	-	-	-
Baseline income	-	-	-	-
Education more than Grade 8	higher*	-	-	-
Married	-	-	-	-
Chronic Depressed	higher*	higher*	higher*	higher*
Growth Rate				
Base Rate	-0.81			
Age	positive*			
Baseline Income	negative*			
Chronic Depressed	negative*			

* denotes statistical significance at .10 or less

1. Base Status of HFSD is different from that of Cal-SAHF. For HFSD, it denotes a base status value of the PSI score at the first follow-up (i.e., an intercept or constant), whereas for Cal-SAHF, the base status represents the PSI scores for those who are Latino-Spanish speaking and have high school education (beyond 8th grade).

2. The term 'lower' or 'higher' indicates that expected score for each predictor (i.e., Latino/Spanish speaking, Asian, Baseline income, etc.) is either lower or higher than Base Status reported at the top of each column.

3. In HLM a base status, representing the mean score of a group, using combined variables, is used to compare the scores of different groups. The base status is determined statistically, using theoretical expectations and groupings in which there is sufficient data to meet statistical requirements. The expected scores for other groupings on the identified variables are then compared to see if the expected values of each differ significantly from the base status.

Table 9
Summary of HLM findings on HOME inventory

	HFSD (n = 338)		Cal-SAHF (n = 120)	
HOME Inventory	Baseline	Year 3	Year 1	Year 2
Initial Status				
Base Status [1,3]	33.7*	37.6*	35.8*	34.8**
Age	higher[2]*	-	lower*	-
Latino/English Speaking	-	-	lower*	-
Latino/Spanish Speaking	lower[2]*	-	-	-
African American	lower*	-	-	-
Asian	-	higher*	lower *	-
White	-	-	-	higher *
Baseline income	-	-	-	-
Intervention	-	-	-	-
Education more than Grade 8	higher*	higher*	-	-
Married	-	-	-	-
Chronic Depressed	lower*	lower*	-	-

* denotes statistical significance at .10 or less

[1] Base Status of HFSD is different from that of Cal-SAHF. For HFSD, it denotes a base status value of the HOME score at the first follow-up (i.e., an intercept or constant), whereas for Cal-SAHF, the base status represents the HOME score for those who are Latino-Spanish speaking and have high school education (beyond 8th grade).

[2] The term 'lower' or 'higher' indicates that expected score for each predictor (i.e., Latino/Spanish speaking, Asian, Baseline income, etc.) is either lower or higher than Base Status reported at the top of each column.

3 In HLM a base status, representing the mean score of a group, using combined variables, is used to compare the scores of different groups. The base status is determined statistically, using theoretical expectations and groupings in which there is sufficient data to meet statistical requirements. The expected scores for other groupings on the identified variables are then compared to see if the expected values of each differ significantly from the base status.

Discussion

The results show that some clients improved with regard to the selected outcome variables. The assumption has been that participating in a family support programme would result in reductions on some of the risk measures (depression, substance abuse, violence). As the analysis progressed depression was considered as a risk factor, and improved depression scores were considered as an outcome. Improvements in the parents are presumed to result in improvements in the child. In this analysis, the sole measure of child outcomes that could be compared across programmes had to do with mental development. It may be that the changes of interest are not being captured by a simple measure of mental development. Nevertheless, some initial improvements in Bayley MDI scores were observed in the HFSD intervention group. These differences were not sustained in the third year. There are three possible explanations for this: 1) the measurement changed between year 2 and 3 from the Bayley MDI to the Stanford- Binet (Landsverk, et.al., 2001); 2) the intervention did not specifically address the child's cognitive development; 3) the intervention was too weak to completely reverse a tendency for high risk children to obtain lower IQ scores over time.

The findings indicate that the HFSD and Cal-SAHF populations were demographically similar, with the ABC sample representing a population that differed in that it was older and married. There were more Hispanic families in the ABC population than in the other two samples. It should be noted that while the HFSD and Cal-SAHF study samples included a similar number of Asians, it is likely that the two studies contained very different Asian groups. This may account for the observation that Asian women in the HFSD sample tended to be more depressed, and that Asian women in the two samples differed in their PSI scores. Unfortunately, the data do not provide insights into which Asian groups were represented in the population, so important information about values, ethnicity, and acculturation as they affect some of the outcomes were not able to be studied. The three samples all represented high risk populations in that there was a high incidence of depression, violence (as measured on the CTS), and substance use at baseline. In the HLM analysis, there is some evidence that ethnicity and language affected the outcomes. Additional interactive factors included education and income. In hindsight, it is recognized that the analysis would have been richer had it been possible to distinguish ethnic and cultural factors in a more refined way. In particular, some of the effects of acculturation might have yielded important explanations for some of the HLM results.

A relationship between depression and intensity of services as measured by the number of service transaction was found in HFSD. While the programmes were intended to operate very similarly, implementation, as measured in direct services delivered, looked very different across the three programmes. Interestingly, all three populations seemed to improve over time, although in the analysis, a clear relationship was not found between the intervention and the improvement.

With respect to key outcomes measures, it was found that a reduced depression for all groups, but also noted that baseline depression exerts a strong influence on the intensity of services and on the changes in the PSI and HOME scores. It was notable that depression, age, ethnicity, and baseline income seem to affect the outcomes measures. It is important to note that for HFSD, being assigned to the intervention group did not affect the outcomes. One explanation for this is that the programme did not specifically address issues of chronic depression, and may not have taken culture sufficiently into account. Nonetheless, the rate of decline in depression was faster for the HFSD intervention group. The HLM analysis indicates some differences between the HFSD intervention and control groups with respect to the *rates* of change, yet the differences in outcomes scores between the intervention and control group were not significant.

Conclusions

In this analysis of three related programmes it was found that families generally improved over time. The findings suggest that future studies might focus on better capturing the dynamics associated with age, income, ethnicity and education. In hindsight, some aspects of culture and ethnicity were not sufficiently measured in these three programmes. For example, measures of class, level of assimilation, and religion might be helpful in understanding the observed interaction between ethnicity and outcomes.

In order to adequately evaluate programme effectiveness, the programme model needs to be focused enough on its outcomes. Clarity about the 'unit of intervention' (parent, child, a dyad of parent-child etc.,) would improve the programme's ability to connect specific activities to specific outcomes (Hernandez, 2000). An important note here is that although these three programmes utilized *home visiting* as an intervention strategy, they may not be evaluations of 'the efficacy of home visiting', primarily because home visiting is a means of delivering a service, not a service in and of itself. Future initiatives would be well served by carefully delineating the components of the services that they are proposing, along with carefully outlining the way in which these services will be delivered (Simmel, 2002; Leff & Mulkern, 2002). Careful content analysis of the three programmes discussed here might indicate that the actual service provided was *clinical case management*, and this service was sometimes provided in the home, and sometimes provided in center-based environments.

Along with more specific models connecting specific interventions to outcomes, it is important to assure that the programme intervention being delivered maintains a high level of implementation fidelity (Harachi, et.al., 1999; Leff & Mulkern, 2002). This means that interventions need to be specifically targeted and defined, and there need to be clearly specified intervention activities associated with each of the

outcomes identified in the logic model. Specific interventions addressing parental depression and child mental and social development need to be clearly embedded in the intervention. Specific evaluation of the level of implementation will help with efficacy studies as well as with effectiveness studies. In order to understand how clients did in the programme, it is important to understand more clearly what their experience in the programme was.

Note

1. Education greater than 8th grade indicates successful completion of primary and middle school levels in the United States educational system, and entry into the next level, high school. Normally this is expected to occur at age 14 or 15.

References

Abidin, R. (1995) *Parenting Stress Index.* Psychological Assessment Resources Inc., 16204 N. Florida Ave., Lutz, Florida, 33549

Bayley, N. (1993) *Bayley Scales of Infant Development* (2nd ed.) New York: The Psychological Corporation

Barlow, J., Stewart-Brown, S., Callaghan, H., Tucker, J., Brocklehurst, N., Davis, H., and Burns, C. (2003) *Child Abuse Review*, 12, 172-189

Bavolek, S.J. and Keene, R.G. (1999) Handbook *for the AAPI-2: Adult-adolescent parenting inventory*. Park City, UT: Family Development Resources

Caldwell B., and Bradley, R. (1984) *Home Observation for Measurement of the Environment, (HOME)* (revised ed.) HOME INVENTORY LLC, Distribution Center, and 2627 Winsor Drive, Eau Claire, WI 54703

Carrilio, T. (1998) *The California Safe and Healthy Families Family Support Home Visiting Model: Executive summary.* (2nd ed.) Sacramento, CA: California Office of Child Abuse Prevention

Carrilio, T. (2001) Family support program development: Integrating research, practice and policy. *Journal of Family Social Work*, 6, 3, 53-78

Carrilio, T.E. (2003) Learning from experience? A review of three California initiatives addressing the needs of vulnerable families. *Social Policy Journal*, 2, 2/3, 5-25

Carrilio, T.E. and Min, J.W. (2003) *Data analysis of Healthy Families-San Diego, CALSAHF, and ABC.* [Final report. Presented to the California Department of Social Services] San Diego, CA: Social Policy Institute

Carrilio, T., Packard, T., Canady, L., Darling, M., Bergthold, T., Chamberlin. J., Doroski, E., and Gamble, L. (2002) Final *Report: Answers benefiting children program evaluation*. [Technical

report presented to the California Department of Social Services Office of Child Abuse Prevention] San Diego, CA: Social Policy Institute

Carrilio, T.E., Packard, T.R., and Clapp, J.D. (2003) Nothing in, nothing out: Barriers to data based program planning. *Administration in Social Work,* 27, 4, 61-75

Daro, D. (2005) Letter to the Editor: Response to Chaffin, 2004. *Child abuse and Neglect*, 29, 237-240

Daro, D. and Donnelly, A.C. (2002) Charting the waves of prevention: Two steps forward, one step back. *Child Abuse and Neglect*, 26, 731-742

Daro, D., McCurdy, Falconnier, L., and Stojanovic, D. (2003) Sustaining new parents in home visitation services: Key participant and program factors. *Child Abuse and Neglect*, 27, 1101-1125

Duggan,A., McFarlane,E., Fuddy,L., Burrell,L., Higman, S.M.,Windham, A., and Sia, C. (2004) Randomized trial of a statewide home visiting program: Impact in preventing child abuse and neglect. *Child Abuse and Neglect*, 28, 597-622

Duggan, A., McFarlane, E., Windham, A., Rohde, C., Salkever, D., Fuddy, L., Rosenberg, L., Buchbinder, S., and Sia, C. (1999) Evaluation of Hawaii's Healthy Start Program. in R. Behrman *The Future of Children: Recent program evaluations*, 9, 1, 66-90. Los Altos, CA: David and Lucille Packard Foundation.

Ewing, J., and Mayfield, D. (1974) CAGE. *American Journal of Psychiatry*, 131, 112-122

Gavin, D.R., Ross. H.E., and Skinner, H.A., (1989) Diagnostic validity of the Drug Abuse Screening Test in the assessment of DSM-III drug disorders. *British Journal of Addiction*, 84, 301-307

Gomby, D. (1999) Understanding evaluations of home visitation programs. in R. Behrman, *The Future of Children: Recent program evaluations*, 9, 1, 27-43. Los Altos, CA: David and Lucille Packard Foundation

Gomby, D., Culross, P., and Behrman, R. (1999) Home visiting: Recent program evaluations – analysis and recommendations. in R. Behrman *The Future of Children: Recent program evaluations*, 9, 1, 4-26. Los Altos, CA: David and Lucille Packard Foundation

Gomby, D., Larson, C., Lewitt, E.M., and Behrman, R. (1993) Home visiting: analysis and recommendations. in R, Behrman *The Future of Children: Home visiting*, 3, 3, 6-22. Los Altos, CA: David and Lucille Packard Foundation

Guterman, N.B. (2001) *Stopping Child Maltreatment before it Starts*. Thousand Oaks, CA: Sage

Hahn, R.A., Mercy, J., Bilukha, O., and Briss, P. (2005) Assessing home visiting programs to prevent child abuse: Taking silver and bronze along with gold. *Child abuse and Neglect*, 29, 3, 215-218

Harachi, T.W., Abbott, R.D., Catalano, R.F., Haggerty, K.P., and Fleming, C.B. (1999) Opening the black box: Using process evaluation measures to assess implementation and theory building.. *American Journal of Community Psychology,* 27, 5, 711-731

Hernandez, M. (2000) Using logic models and program theory to build outcome accountability. *Education and Treatment of Children*, 23, 1, 24-40

Illig, D. (1998) *Birth to Kindergarten: The importance of the early years.* Sacramento: California

Research Bureau, California State Library CR-98-001

Karoly, L. Greenwood, P., Evringham, S.Hoube, J., Kilburn, M.R., Rydell, P., Saunders, M., and Chiesa, J. (1998) *Investing in our Children: What we know and don't know about the costs and benefits of early childhood interventions.* Santa Monica, CA: Rand Corporation, Rand MR 898-TCWF.

Kitzman, H., Olds D.L., Henderson, C.R. Jr, Hanks, C., Cole, R., Tatelbaum, R., McConnochie, K.M., Sidora, K., Luckey, D.W., Shaver, D., Engelhardt K., James, D., and Barnard, K. (1997) Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing. A randomized controlled trial. *Journal of the American Medical Association*, 278, 8, 644-52

Krugman, R. (1993) Universal home visiting: A recommendation from the U.S. Advisory Board on Child Abuse and Neglect. in R. Behrman *The Future of Children: Home visiting*, 3, 3, 184-191.Los Altos, CA: David and Lucille Packard Foundation

Landsverk, J., Carrilio, T., Connelly, C., and Ganger, W. (2001) *Healthy Families San Diego. [*Final technical report, submitted to the California Department of Social Services, The Wellness Foundation, and Stuart Foundation] San Diego, CA: Child and Adolescent Services Research Center

Leff, S.H. and Mulkern, V. (2002) Lessons learned about science and participation from multisite evaluations. in J.M. Herrell, and R.B. Straw *New Directions for Evaluation*: *Conducting multiple site evaluations in real-world settings*. San Francisco: Jossey Bass

McGuigan, W.M., Katzev, A.R., and Pratt, C. (2003) Multi-level determinants of retention in a home-visiting child abuse prevention program. *Child Abuse and Neglect*, 27, 363-380

Olds, D. (2003) Reducing program attrition in home visiting: What do we need to know? *Child Abuse and Neglect*. 27, 359-361

Olds, D.. Eckenrode, J.. Henderson, C.. Kitzman, H.. Powers, J., Cole, R., Sidora, K., Morris, P., Pettitt, L., and Luckey, D.(1997) Long-term effects of home visitation on maternal life course and child abuse and neglect: Fifteen-year followup of a randomized trial. *Journal of the American Medical Association*, 278, 8, 637-643

Olds, D., Henderson, C., Kitzman, H., Eckenrode, J., Cole, R., and Tattlebaum, R. (1999) Prenatal and infancy home visitation by nurses. in R. Behrman *The Future of Children: Recent program evaluations*, 9, 1,44-63. Los Altos, CA: David and Lucille Packard Foundation

Patton, M. (1997) *Utilization-focused Evaluation*. Thousand Oaks, CA: Sage

Radloff, L.S. (1977) The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401

Raudenbush, S.W., and Bryk, A.S. (2002) *Hierarchical Linear Models: Applications and data analysis methods.* (2nd Ed.). Thousand Oaks: Sage

Reynolds, A.J., Temple, J., Robertson, D., and Mann, E. (2001) Long term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association*, 285, 18, 2339-2346

Saunders, J.B., Asland, O.G., Babor, T.F., de la Fuente, J.R., and Grant, M. (1993) Development of The Alcohol Use Disorders Identification Test (AUDIT). *WHO Collaborative Project on*

Early Detection of Persons with Harmful Alcohol Consumption. II. Addiction, 88, 791-804

Schorr, L. (1997) *Common Purpose*. New York: Anchor Doubleday

Sherwood, K.E. (2005) Evaluating home visitation: A case study of evaluation at the David and Lucile Packard Foundation. in M.Q. Patton, and P. Patrizi *Teaching Evaluation Using the Case Method. New directions for evaluation.* No. 105. pp. 59-78. San Fancisco: Jossey-Bass

Shonkoff, J.P., and Phillips, D.A. (2000) *From Neurons to Neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press

Simmel, C. (2002) The shared family care development project: Challenges of implementing and evaluating a community based project. *Children and Youth Services Review*, 24, 6-7, 455-470

Straus, M.A., Hamby, S.L., Boney-McCoy, S., and Sugarman, D.B. (1996) The Revised Conflict Tactics Scales (CTS2). *Journal of Family Issues*, 17, 3, 283-316

Straw, R.B. and Herrell, J.M. (2002) A framework for understanding and improving multisite evaluations. in J.M. Herrell, and R.B. Straw *New Directions for Evaluation: Conducting multiple site evaluations in real-world settings*, 94, pp.5-15. San Francisco: Jossey Bass

Sweet, M.A. and Appelbaum, M.I. (2004) Is Home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development*, 75, 5, 1435-1456

Wagner, M., Spiker, D., Linn, M.I., Gerlach-Downie, S., and Hernandez, F. (2003) Dimensions of parental engagement in home visiting programs: Exploratory study. *Topics in Early Childhood Special Education*, 23, 4, 171-187

Wasik, B., Bryant, D., Sparling, J., and Ramey, C. (1997). Mediating variables: Maternal problem solving. in R.T. Gross, D. Spiker, and C. Haynes *The Infant Health and Development Program*. Stanford, CA: Stanford University Press

Whyte, W., ed. (1991) *Participatory Action Research*. Newbury Park, CA: Sage